```
   1 CTCTGCAGCT CAGCATGGCT AGGGTACTGG GAGCACCCGT TGCACTGGGG
  51 TTGTGGAGCC TATGCTGGTC TCTGGCCATT GCCACCCCTC TTCCTCCGAC
 101 TAGTGCCCAT GGGAATGTTG CTGAAGGCGA GACCAAGCCA GACCCAGACG
 151 TGACTGAACG CTGCTCAGAT GGCTGGAGCT TTGATGCTAC CACCCTGGAT
 201 GACAATGGAA CCATGCTGTT TTTTAAAGGG GAGTTTGTGT GGAAGAGTCA
 251 CAAATGGGAC CGGGAGTTAA TCTCAGAGAG ATGGAAGAAT TTCCCCAGCC
 301 CTGTGGATGC TGCATTCCGT CAAGGTCACA ACAGTGTCTT TCTGATCAAG
 351 GGGGACAAAG TCTGGGTATA CCCTCCTGAA AAGAAGGAGA AAGGATACCC
 401 AAAGTTGCTC CAAGATGAAT TTCCTGGAAT CCCATCCCCA CTGGATGCAG
 451 CTGTGGAATG TCACCGTGGA GAATGTCAAG CTGAAGGCGT CCTCTTCTTC
 501 CAAGGCCATG GACACAGGAA TGGGACTGGC CATGGGAACA GTACCCACCA
 551 TGGCCCTGAG TATATGCGCT GTAGCCCACA TCTAGTCTTG TCTGCACTGA
 601 CGTCTGACAA CCATGGTGCC ACCTATGCCT TCAGTGGGAC CCACTACTGG
 651 CGTCTGGACA CCAGCCGGGA TGGCTGGCAT AGCTGGCCCA TTGCTCATCA
 701 GTGGCCCCAG GGTCCTTCAG CAGTGGATGC TGCCTTTTCC TGGGAAGAAA
 751 AACTCTATCT GGTCCAGGGC ACCCAGGTAT ATGTCTTCCT GACAAAGGGA
 801 GGCTATACCC TAGTAAGCGG TTATCCGAAG CGGCTGGAGA AGGAAGTCGG
 851 GACCCCTCAT GGGATTATCC TGGACTCTGT GGATGCGGCC TTTATCTGCC
 901 CTGGGTCTTC TCGGCTCCAT ATCATGGCAG GACGGCGGCT GTGGTGGCTG
 951 GACCTGAAGT CAGGAGCCCA AGCCACGTGG ACAGAGCTTC CTTGGCCCCA
1001 TGAGAAGGTA GACGGAGCCT TGTGTATGGA AAAGTCCCTT GGCCCTAACT
1051 CATGTTCCGC CAATGGTCCC GGCTTGTACC TCATCCATGG TCCCAATTTG
1101 TACTGCTACA GTGATGTGGA GAAACTGAAT GCAGCCAAGG CCCTTCCGCA
1151 ACCCCAGAAT GTGACCAGTC TCCTGGGCTG CACTCACTGA GGGGCCTTCT
1201 GACATGAGTC TGGCCTGGCC CCACCTCCTA GTTCCTCATA ATAAAGACAG
1251 ATTGCTTCTT CGCTTCTCAC TGAGGGGCCT TCTGACATGA GTCTGGCCTG
1301 GCCCCACCTC CCCAGTTTCT CATAATAAAG ACAGATTGCT TCTTCACTTG
1351 AATCAAGGGA CCTTGGTCGT GAAACAATCT TCTTTCTTTG AGTTGAAAAG
1401 TTAGCACTTC TCCTTTGAGG GTGTCGAGCT CAAACAAGGC TGTGAGAAAC
1451 AAGGGAGGGG AGCACTAAGG GGCAAACCTA TCTCTGCGCA GATGATTCTT
1501 AGGTCCAGAT CATAAACTAG CTCTTTGCAG ACTATCTACA CATAGTGGGG
1551 GGAAAGAGAA CCAGAGTCGG AAGAGGAACA GCTGAGTTTA TACAGCAAGT
1601 AAGAGGTGGA GCTAGGACTC TGATTCAACT TGCTGGTAGA TGGCCACAAC
1651 CCAGCCGCAA GGCATCAGAA ACAACAGGGC CTGGGGCAAC TATGCATGTG
1701 CAAAGAGGAT TGGCTCAGAG TTGTGGGGTA GGAGGTCCAA TCTGGGGGAC
1751 CTCAAATTAT GGTTCTGGGT GATTCAAGTA ACACCACTCA TGGCTTGTGT
1801 TGCCATGAGT TAGGCATGAC AAGTGGAATG AAGTTGAAGT GGGGAAACAG
1851 AAATACACCA GCTGTGTGTC AGAGGCAAGC TGGAGAGAGA GAAGAAAGAA
1901 TGAATGGCAC CATGGAGCAC ATTTGCAGAA CACAGTCCCT GGGAGTCTTG
1951 CTGGAGCCTC AGGAGCTTTG CTGGCACAGA GGATCTGGCC TACCCAATTA
2001 GCCTCCTGGG TATCTGCACC ATCTAGACCA GCAAATGTCA CTGGCAAGGA
2051 GGTTGCAGTG CTTGGTTATT TTCTGGTCAT AAACTGGTGA AGGCTTTGGG
2101 TTCCAAATTT GCTGACAGCT GTTTAACTGG GAATTGGGCC TAGACTATAG
2151 GTAGCTATGT CTCAGACAAG GCCCTATTCC TCCACTGCCT TTACAACCCA
2201 GCTGAGGTTG GAGGCTGGCT TGTTTCAGCC TCAAAAAATA GCCTGAGTTT
2251 CCAGCAGAGG GCCCTTATTC TGAGCTTCCG TGTCCTAGCC TCATTTTCCT
2301 TTCCTGTAAA ATAGACACAA TGCCACCCAC CTTCCAGTGA CAATGAATAT
2351 AGACTCAAAC CCATCCCTTG AACTGTCTTG GGAAGGGGCT CTGGACGTAG
2401 ACCCAGACTG TGGCTCATGG CCTCATGTGA TCTGGAGTCA GCCCCTCCCA
2451 ACCTGTCAGC CATTTGCTCC GTAGGACTTT GATGGGTAGA GTAGTAGCTA
2501 ACAAGCTCTG ACTGTCACAC AAGGCTTTGT ACTGGGAGGC CAGGCTATAG
2551 AGTGGCTCCA GCTTAAAGGG CTGGGAGCTG GGGGACAGTG TCTCAGATTA
2601 GGGTCTAACT AGGAAGTTGA CTGGAGCTGA GAACAGAGGT TAGGGGCCAA
2651 GCAGCAGGGT TGTGGGTCTA CTCCTTAGGA GCACCTTGAG CTTTACTTTT
2701 CATTCCTAAT GGTGTCTTGG ATGGCTACCC TCACGGGGTT GGCTGCTAGT
2751 CTAAGGGGTG GAGACAAGGA CAGAGTTTCA GGTCTGGTCC TTATCAAGTT
2801 CATGCACTAC ACTTGGGACC ACTGCTGCAT CATGCCAGGG AGCCTAGAGG
2851 TGTCTAAACA GTTATCCAAC AACTGTGATA CCCAAGGTTA ACTTTCTCTT
2901 GTTTTCAGAG GCAGGGAGTA CTAAGTCTCC CCTTTCTCCT TTCCTCCCAC
2951 GTGTTCTCTT GCAGGGAATC CTCTAGCTTG TCTCCAGGGA ACTCCCAGAA
3001 ATGGTTTGTT TCAGTCAGTT TAGGCTGCTA TAAGAGAATA TCTTAGAGTG
3051 GGTAATCTAT CAGCAATAGG AATTTATTGT TCACAATTCT GGAGGCTGGA
3101 AAATCCAAGA TCAAGGCTCC AGCAGGTTCA GTGTCTGCTG AGTGCTTGTT
```

FIGURE 1, page 1 of 2

3151 CTGCTTCGAA GATGGCACCT TTTTGCTGTG TTCTCA (SEQ ID NO: 1)

FEATURES:

5'UTR:	1-14
Start Codon:	15
Stop Codon:	1188
3'UTR:	1191

Homologous proteins:

Top 10 BLAST Hits:

	Score	E
CRA|335001098638983 /altid=gi|11321561 /def=ref|NP_000604.1| he...	681	0.0
CRA|18000004928118 /altid=gi|386789 /def=gb|AAA52704.1| (J03048...	679	0.0
CRA|18000005034645 /altid=gi|1335098 /def=emb|CAA26382.1| (X025...	634	0.0
CRA|18000004885233 /altid=gi|1708184 /def=sp|P20058|HEMO_RABIT ...	519	e-146
CRA|18000004905757 /altid=gi|1070649 /def=pir||OQRB hemopexin p...	513	e-144
CRA|84000015361878 /altid=gi|13641048 /def=ref|XP_011963.2| hem...	504	e-141
CRA|18000004936853 /altid=gi|123036 /def=sp|P20059|HEMO_RAT HEM...	466	e-130
CRA|18000004882890 /altid=gi|1708183 /def=sp|P50828|HEMO_PIG HE...	459	e-128
CRA|18000005011238 /altid=gi|1087020 /def=gb|AAA82488.1| hepato...	436	e-121
CRA|18000005041763 /altid=gi|1311343 /def=pdb|1HXN| Heme Mol...	408	e-113

Blast hits to dbEST:

	Score	E
gi|12798347 /dataset=dbest /taxon=960...	1360	0.0
gi|12914625 /dataset=dbest /taxon=960...	1344	0.0
gi|6360478 /dataset=dbest /taxon=9606 ...	973	0.0
gi|9866417 /dataset=dbest /taxon=960...	967	0.0
gi|12798348 /dataset=dbest /taxon=960...	839	0.0

Expression Information:

Tissue source of BLAST dbEST hits:

gi|12798347 Fetal brain
gi|12914625 brain neuroblastoma cells
gi|6360478 liver
gi|9866417 non cancerous liver tissue
gi|12798348 Fetal brain

Tissue source of cDNA clone:

Fetal liver

FIGURE 1, page 2 of 2

```
   1 MARVLGAPVA LGLWSLCWSL AIATPLPPTS AHGNVAEGET KPDPDVTERC
  51 SDGWSFDATT LDDNGTMLFF KGEFVWKSHK WDRELISERW KNFPSPVDAA
 101 FRQGHNSVFL IKGDKVWVYP PEKKEKGYPK LLQDEFPGIP SPLDAAVECH
 151 RGECQAEGVL FFQGHGHRNG TGHGNSTHHG PEYMRCSPHL VLSALTSDNH
 201 GATYAFSGTH YWRLDTSRDG WHSWPIAHQW PQGPSAVDAA FSWEEKLYLV
 251 QGTQVYVFLT KGGYTLVSGY PKRLEKEVGT PHGIILDSVD AAFICPGSSR
 301 LHIMAGRRLW WLDLKSGAQA TWTELPWPHE KVDGALCMEK SLGPNSCSAN
 351 GPGLYLIHGP NLYCYSDVEK LNAAKALPQP QNVTSLLGCT H  (SEQ ID NO:2)
```

FEATURES:
Functional domains and key regions:
Prosite results:

```
PDOC00001 PS00001 ASN_GLYCOSYLATION
N-glycosylation site
Number of matches: 4
       1      64-67   NGTM
       2      169-172        NGTG
       3      175-178        NSTH
       4      382-385        NVTS

PDOC00005 PS00005 PKC_PHOSPHO_SITE
Protein kinase C phosphorylation site
Number of matches: 5
       1      47-49   TER
       2      78-80   SHK
       3      87-89   SER
       4      216-218        TSR
       5      298-300        SSR

PDOC00006 PS00006 CK2_PHOSPHO_SITE
Casein kinase II phosphorylation site
Number of matches: 9
       1      40-43   TKPD
       2      59-62   TTLD
       3      95-98   SPVD
       4      141-144        SPLD
       5      216-219        TSRD
       6      235-238        SAVD
       7      242-245        SWEE
       8      321-324        TWTE
       9      366-369        SDVE

PDOC00008 PS00008 MYRISTYL
N-myristoylation site
Number of matches: 5
       1      6-11    GAPVAL
       2      170-175        GTGHGN
       3      201-206        GATYAF
       4      279-284        GTPHGI
       5      317-322        GAQATW

PDOC00009 PS00009 AMIDATION
Amidation site
              305-308        AGRR

PDOC00013 PS00013 PROKAR_LIPOPROTEIN
Prokaryotic membrane lipoprotein lipid attachment site
              379-389        QPQNVTSLLGC
```

PDOC00023 PS00024 HEMOPEXIN
Hemopexin domain signature
Number of matches: 2

1	86-101	ISERWKNFPSPVDAAF
2	226-241	IAHQWPQGPSAVDAAF

Membrane spanning structure and domains:

Helix	Begin	End	Score	Certainity
1	6	26	1.820	Certain
2	251	271	0.639	Putative

FIGURE 2, page 2 of 3

BLAST Alignment to Top Hit:

```
>CRA|335001098638983 /altid=gi|11321561 /def=ref|NP_000604.1|
          hemopexin [Homo sapiens] /org=Homo sapiens /taxon=9606
          /dataset=nraa /length=462
         Length = 462

 Score =  681 bits (1737), Expect = 0.0
 Identities = 341/468 (72%), Positives = 351/468 (74%), Gaps = 83/468 (17%)

Query: 1   MARVLGAPVALGLWSLCWSLAIATPLPPTSAHGNVAEGETKPDPDVTERCSDGWSFDATT 60
           MARVLGAPVALGLWSLCWSLAIATPLPPTSAHGNVAEGETKPDPDVTERCSDGWSFDATT
Sbjct: 1   MARVLGAPVALGLWSLCWSLAIATPLPPTSAHGNVAEGETKPDPDVTERCSDGWSFDATT 60

Query: 61  LDDNGTMLFFKGEFVWKSHKWDRELISERWKNF---------------------------- 93
           LDDNGTMLFFKGEFVWKSHKWDRELISERWKNF
Sbjct: 61  LDDNGTMLFFKGEFVWKSHKWDRELISERWKNFPSPVDAAFRQGHNSVFLIKGDKVWVYP 120

Query: 94  --------------------PSPVDAAFR--QGH---NSVFLIKGDKVWVYP---PEKKEK 126
                               PSP+DAA    +G       V   +GD+ W +        KE+
Sbjct: 121 PEKKEKGYPKLLQDEFPGIPSPLDAAVECHRGECQAEGVLFFQGDREWFWDLATGTMKER 180

Query: 127 GYPK----------------------LLQDEFPG-IPSPLDAAVECHRGECQAEGVLFFQ 163
            +P                        L  D   G +P      V  +   C         +
Sbjct: 181 SWPAVGNCSSALRWLGRYYCFQGNQFLRFDPVRGEVPPRYPRDVRDYFMPCPG------R 234

Query: 164 GHGHRNGTGHGNSTHHGPEYMRCSPHLVLSALTSDNHGATYAFSGTHYWRLDTSRDGWHS 223
           GHGHRNGTGHGNSTHHGPEYMRCSPHLVLSALTSDNHGATYAFSGTHYWRLDTSRDGWHS
Sbjct: 235 GHGHRNGTGHGNSTHHGPEYMRCSPHLVLSALTSDNHGATYAFSGTHYWRLDTSRDGWHS 294

Query: 224 WPIAHQWPQGPSAVDAAFSWEEKLYLVQGTQVYVFLTKGGYTLVSGYPKRLEKEVGTPHG 283
           WPIAHQWPQGPSAVDAAFSWEEKLYLVQGTQVYVFLTKGGYTLVSGYPKRLEKEVGTPHG
Sbjct: 295 WPIAHQWPQGPSAVDAAFSWEEKLYLVQGTQVYVFLTKGGYTLVSGYPKRLEKEVGTPHG 354

Query: 284 IILDSVDAAFICPGSSRLHIMAGRRLWWLDLKSGAQATWTELPWPHEKVDGALCMEKSLG 343
           IILDSVDAAFICPGSSRLHIMAGRRLWWLDLKSGAQATWTELPWPHEKVDGALCMEKSLG
Sbjct: 355 IILDSVDAAFICPGSSRLHIMAGRRLWWLDLKSGAQATWTELPWPHEKVDGALCMEKSLG 414

Query: 344 PNSCSANGPGLYLIHGPNLYCYSDVEKLNAAKALPQPQNVTSLLGCTH 391
           PNSCSANGPGLYLIHGPNLYCYSDVEKLNAAKALPQPQNVTSLLGCTH
Sbjct: 415 PNSCSANGPGLYLIHGPNLYCYSDVEKLNAAKALPQPQNVTSLLGCTH 462  (SEQ ID NO:4)
```

Hmmer search results (Pfam):

Scores for sequence family classification (score includes all domains):

Model	Description	Score	E-value	N
PF00045	Hemopexin	154.0	3.1e-45	4
CE00423	E00423 stromelysin_1	10.9	0.014	2

Parsed for domains:

Model	Domain	seq-f	seq-t		hmm-f	hmm-t		score	E-value
PF00045	1/4	56	95	..	1	50	[]	25.3	4.5e-06
CE00423	1/2	40	117	..	294	375	..	11.5	0.0096
PF00045	2/4	97	141	..	1	50	[]	58.1	4.7e-16
PF00045	3/4	192	235	..	1	50	[]	37.9	6.5e-10
CE00423	2/2	206	241	..	323	358	..	2.2	4.7
PF00045	4/4	237	282	..	1	50	[]	35.0	4.8e-09

FIGURE 2, page 3 of 3

1 TCCCTCTCCC CAGGCAGGCC CAGCAAAATC TGTAGGATTC AGACAGGGTT
51 CTGACAGCTG AAGACAAGTT GTTGAGGAAA TTCCTGATGG AGGATCATGG
101 GGTGCTCAGG AGGGAGAATA TAAGGTTTCA GAGGCTGAGA GGGAAAGAAA
151 AGGTGAGGGG GAGTCTTAGA ATAGTGGCTC CCATTGCCCA ACACCCAGAA
201 AGAAGACATG CCCTGCAATG GGGAGAAGGT GAGTATGAGA CATTGGCTGT
251 AGCAGCGATG GCATTGCCCA GGCTGCCAAG GACTCAGAGA GTCCAGCCTT
301 GCCCACTGAC CTATGAGGAG GGAATGATGT TCACAGCACA TTTTCATTCG
351 TAAGTCAGGA GAGGACATTG AGCCTGATGG CAGAGGCCTG GTGACATGTT
401 GTTCCAGAGG TTCCGGAATG TGTGTTTTCC TGTTGGAAGG AAACTTCGCA
451 GAGTAGAAAA GGGATCTGAG ACTTTTGGTA AGATTATATA TGGGACTGTC
501 AGGGGTCTGG AGCCATCTGT GAGGGATCAG GGCCCTTTCA GCCTTGGCTA
551 GGGAGCAGGG GTCCTGGAAC TTCATCCTGG CCCATAGCTG AGTCTGCCCA
601 TAATTCTTTT CTGACTCACT AGGCAAATCT CACACAGAAA TGGGGCAGCT
651 TTGGGAGTGG GCCCAGGAAG TACTGAGGAT AGCAGGTGAG ATCCCAGGAA
701 GAGATGGATG TGGGGCCGAG ACACTGGAGA GAGAAACAGG ACTGTCAGAT
751 AAAGGGCGTC TGTGACTCCT AGATCTCATT ATGCCTACTA CCATAACCTA
801 CCCCCAATTC CTAATATTCT CCTACCCTAG AGGGGGGGAA ATTGTCAGAA
851 ATTTGGCTGC AACACTAGCA ACACTACTCA GTACTTGAAA TGCATTTTTG
901 CATTTTTTTC ATTCAACAAA TATTTCTGGA ACAACTCTTA TATGCCAGGC
951 ACTATTTTAG GAGTCAGGGA TATATAATGG TAAACAAGAC AGGCAAAACA
1001 AAGCAAAGCA ACAACAACCA TCACCAGATA AGTAGACAGA TGAAAGAATT
1051 TCAAGTTTTA GTAAGTAAAA TAAAACAAGC AAGGGTCTGA AATGGCTAGA
1101 TAAGGCGGTC AAGAAAGGCT TCATTGAGAA GGTAGCATTT AAGCAGGAGT
1151 CAGCTAGAAA TATTGTGAAA TTCCAGTTAC AGTTCTATTT GTTCTGGGTT
1201 GGTTAAATAA AGCTTTTTCC CCCAAGGTGG AAACTACCAA GAAAGACTAA
1251 TTACTAGTAG TGGTGGTGCT CTCTGGAAGA GAGACACCTC CTGTTTCTGC
1301 CTCATTACTG TCAACCCTTC ACTTCCAGGC ACTTTTTGCA AAGCCCTTTG
1351 CCAGTCAGGG AAGGCGAGAG GCTGGGCATG GGGCTTGGAC ATTTGACAAC
1401 AGTGAGACAT TATTGTCCCC AGACTCACTA GCCCAAGGGT AAAGCTGAAG
1451 AGGCTTGGGC ATGCCCCAGA AAGGCCCCTG ATGAAGCTTG GAAAAAGCTG
1501 TTCTCTGAGT ATTTCTAAGT AAGTTTATCT GTGTGTGTGG TTACTAAAAG
1551 TAGTAAGTAT TGCTGTCTCT AGCTGCCTTA GAGCAGGGCT TGACACAGTA
1601 CACAGCAATA TTAGTTCCCT CCTTTTCTCA CCTCCCCCAT TGTGGAGATA
1651 AACTCAATCA CAAAAGGTGA TCCTCAGTCT ACTCACTTCC CTGACTTATG
1701 GATGCCTGGA CCCATTGCCA GTGTGAGAGT CACAGCTGGA CGTCAGCAGT
1751 GTAGCCCAGT TACTGCTTGA AAATTGCTGA AGGGGGTTGG GGGGCAGCTG
1801 CCGGGAAAAA GGAGTCTTGG ATTCAGATTT CTGTCCAGAC CCTGACCTTA
1851 TTTGCAGTGA TGTAATCAGC CAATATTGGC TTAGTCCTGG GAGACAGCAC
1901 ATTCCCAGTA GAGTTGGAGG TGGGGGTGGT GCTGCTGCCA ACTCTATATA
1951 GGGAGTTCAA CTGGTCACCC AGAGCTGTCC TGTGGCCTCT GCAGCTCAGC
2001 ATGGCTAGGG TACTGGGAGC ACCCGTTGCA CTGGGGTTGT GGAGCCTATG
2051 CTGGTCTCTG GCCATTGCCA CCCCTCTTCC TCCGTGAGTA AAGCTGGGAC
2101 TAGAAGCGAA GGATTGAGTT CTGGGCTAGG GTAAGGTAGG GCCAGTTTTT
2151 AGGCCTCGGT CAAATTTGGG GTCAGGGGCT ATGGGAAAGG GATCGGTCCC
2201 AATGGATCAA GATATCTATT TTGTTCTCCC TAGGACTAGT GCCCATGGGA
2251 ATGTTGCTGA AGGCGAGACC AAGCCAGACC CAGACGTGAC TGGTGAGGCC
2301 CTGACTCCCT AAGTCTGTCT TATCTGTCTG GTTGTGTCTC TGCATTTTAT
2351 CACCTTCTGG TTTTTTTTTT TTTTTTTTTT TTTTACTTTG CCATCTCCCT
2401 ACCTCCACCC CAGAACGCTG CTCAGATGGC TGGAGCTTTG ATGCTACCAC
2451 CCTGGATGAC AATGGAACCA TGCTGTTTTT TAAAGGTAGG AGGGACTGAG
2501 GTTAGGGCGT TTAGGACCTT AGACTTACTC TCCTTCACAA AGGGTGTCCC
2551 TGTCTGTGGG AGGTCTTAGG AATTATCTGA TGGTATCACT GACAGCTTCT
2601 CTCAAGCTAT CTCAGTAGGT CAAAGGTTTC TCACTGGGCC CCTCAGTGAG
2651 TGTGGGTTTT TTCAGGGGAG TTTGTGTGGA AGAGTCACAA ATGGGACCGG
2701 GAGTTAATCT CAGAGAGATG GAAGAATTTC CCCAGCCCTG TGGATGCTGC
2751 ATTCCGTCAA GGTCACAACA GTGTCTTTCT GATCAAGGTA CTGCTGGGCC
2801 AAAATCAGGG CCAGGCTGGA AAGGGCTGGA ATCGACACTG GGGACCCTTC
2851 CCCCAAATGG CCTTGGCATG GAGCCCATAG CAATAGGTAG CAGATTTCTT
2901 TCCCATGTGC CCTCCTTTCC TGTAAAAGCT TGGGCTAAGG GAGTGTGCAT
2951 GCGTGTGGGC CTGGCAGGTG CACCATCCAG TGGCTGTTCT TCAGTCCTAG
3001 TCTTAGTTCT ACACCGCTCT GCTGTACCTC ACACTGCTGG CCATCCTTTT
3051 TTTCTCTGGC AATTGCTTCC CTTGCCTTCC ATGACCCTGT ATCAAGTCCT
3101 CTTCATAGGG CAAGGCAAGT TGTTCCCAAC ACAATGGCAC CTGGCTAGAA

FIGURE 3, page 1 of 8

```
3151 GAGCATGTGG AGCATGAAAT CCAGTCTGCT GTGCTCACCA AGTCCCATGT
3201 GACCCAGGCT GTGTCTGCTC AGAGGAAGGG GTGCCTTTTC CTACCTTGCC
3251 AAAGGTGCTG TGTGGTTGGG GAAGTCCTGA CTGTCGGCTT TGTTTTCCCT
3301 CCTGCCTCTT TTCTCTCTCT TCTCAAATGT CTCATTCTAT CTCAACCAGT
3351 TCCCTAATGT TCCTTGGGGA TCCATCCTAG CCTTTCCATA TACCTTCCCT
3401 CAGTGATCTC AACCATCACC TTGGCTCTGA GGAATATCTA TGCTGTGGAC
3451 ACTGGATCTA GATCTACTTT CTGAGCTCCA GACATCTCTT TCCAATTGTA
3501 TGTTCTACAG GCACCTAAAA TTCAGCATCC CCCAAACTAA GCTTTGCATC
3551 TTCTTTACAA ACCAACCTTT CCTCCTGTGT TTCCTGTTTC AGTAAATGAC
3601 CCCAAAATGT GCCTGATTAC TACAAACCAA GTGCACACAG GGTCTCATGA
3651 TCTGGGCCTT GGTTATCTTC TCAGGTTTAT CTCCTCCCCT GCCACATTCA
3701 CTGTGTGCCA GCCATACGAA TCTACATGAG GTTGGAGCAC ACTGCTTCCT
3751 CATGTTTGGG CTCTGCATGC TGCTCCCTCT GCTGGTAACA CCCTTTCCTC
3801 ACTTGTCAAC CTGGAAAATT CCTGCTGATT TTTCAGCTCT TGGGCCCAAT
3851 GCTTCCTCTT TGGTGTGAAA CCTTCCACAA CTTCTCTAGG CAGACTTAGG
3901 CACTCTGTCT ATATTCTCAG TGCACTCTTT ACACTACACC TTGGTAGTTG
3951 CATGGCTAGG ATTGCAGGAG TCCTTTCTGC TTTTGTACAG TGAACTTCCT
4001 GAAGTGAAAG ACAGAGTCTT GTTATCCTCA GTGCCTCTCA CAATGCCTGG
4051 CATATAGTAG TTATTCAGTG ACTGTTTCTT GGATGAATGA ATGAATGAAT
4101 AAATAAATGA AGAAATGAAT GAAGAAATAA CGTATGGGTG ATTGCAGGAT
4151 GAACAGTTGT GGATATGTTT GTCAACACTG ATAGTGTTGC AGATAAATGT
4201 GCCACAGGAG TGTCTGGGTA CAGAGCTAGA GGCATGTGTG TTATAGTAAT
4251 AGTGACTGGA TTTGCACAAA CTGAGAGTGT GTAATGTGCA AAAGGACAGC
4301 ACATTGTTGT CCACAGATGG ACTGAGAATG TGTAGGGCCA CAGAAGGATA
4351 TCGTATAAGC ACAGTAGATA AAAAATGTGT GTAAATGCAG AGTGGCAGTA
4401 TCTGGGGATG CACAGTCAAA AAGAGAGTAC TTTTGAATGC AGGGGGACAA
4451 AGTCTGGGTA TACCCTCCTG AAAAGAAGGA GAAAGGATAC CCAAAGTTGC
4501 TCCAAGATGA ATTTCCTGGA ATCCCATCCC CACTGGATGC AGCTGTGGAA
4551 TGTCACCGTG GAGAATGTCA AGCTGAAGGC GTCCTCTTCT TCCAAGGTCA
4601 GTCCAGGCTG GAATCCAAGA ACCTGGAGTA GTGGTGGGTT GGTAGTGATG
4651 CCAGTAGTGA TGGTGATAGT GGTAGTGATG GTGGTGGTGG AGCCACTATG
4701 TGGCTTTTTA AGGAAGGGAA ATAGAGAAGC CACGTATGGT CTAGAGGTCA
4751 CGTGAGGGAA GGAGAGGAAG TCATTCTGGT GAAGGCAACT GTGTGTAATT
4801 CTGTGTGAAT AGTCCCTCAT TGTTCCCCAT GACCCTTAGG ACAAATCTAC
4851 CCTCTTTAGT CTTACATACA AGTCTCTCCA TGGCCAAATC CCTATTGGCC
4901 CTTCAGCTTT GACTTTTATT ATACTTTTAC CTTAACACTA AGCTCCAGAA
4951 ACCCTATGCT ATTCTCTGTA CACTCAGTTT GCTCCATGCT TTGGAATCTT
5001 TCCTCTCTCT GGGGTTCCAT CTCTCCTTGT GTGCCTTTTA ATTCCTACTT
5051 CAGATTTCAC TTTAAGTATC ATCTTCCCTG GGAAGTTTTC CCAGACTCTC
5101 CCCACTGCCT TTGCTGAGCT GATCCTGTGT GTTTTGCTGC TGAATTTTGG
5151 TGTATGATCA CCCTCCTTTA GCCATCTCTC TGATGGCTGT GAGCTCCATG
5201 TGGTCAGTAC CATTATCTGG CCCATCCTGG GACCCAGAGA AAGCACAAAG
5251 GAGGGCGTAA CCCGGTCTCA CCAAATGCCT GTTGATTGAT TGGACAAAGG
5301 TGACCGCGAG TGGTTCTGGG ACTTGGCTAC GGGAACCATG AAGGAGCGTT
5351 CCTGGCCAGC TGTTGGGAAC TGCTCCTCTG CCCTGAGATG GCTGGGCCGC
5401 TACTACTGCT TCCAGGGTAA CCAATTCCTG CGCTTCGACC CTGTCAGGGG
5451 AGAGGTGCCT CCCAGGTACC CGCGGGATGT CCGAGACTAC TTCATGCCCT
5501 GCCCTGGCAG AGGTGAGAAA GCCCTAGCAC TTGAGACCTG TCAGAATTCA
5551 TCCACTTTCC CTGAGCTTGT GGATCTCACG TGTCCTAGCT CTCACTTTAA
5601 CTCCGTGTTG CGACACCTTG GCCCTTAATC TAGCCCCATT TCCATTCTGG
5651 ATTTTCCCAT TGCCCTCATA TGGGGAAACC CACACCCCAC TAACCCCAGC
5701 CATCTCTTCC ACCTTGGACC TCACTCTGAC CTCTGGCCTC CTTCTGTGTT
5751 CTCCTCACCC ATTTCTCTCT CCAGGCCATG GACACAGGAA TGGGACTGGC
5801 CATGGGAACA GTACCCACCA TGGCCCTGAG TATATGCGCT GTAGCCCACA
5851 TCTAGTCTTG TCTGCACTGA CGTCTGACAA CCATGGTGCC ACCTATGCCT
5901 TCAGTGGTGA GAGATGCCCC CAACTCCCCC AATGTGCTCT CACATCTCTT
5951 TTACTTGTAT CTCCCATCCT TGACACATTT CTCCATTGTC ATCACTGTGT
6001 CACTTATTTT GTCCCCTCTG TCCCCATCCT TCTGCATGCC CTTCTGCATC
6051 CCTCATCTCT GAGGCATATT TCTCAATCTT GTCTGTCACG GCCCAAGCCC
6101 CTAACTTCAT CTACCTGTCT ACCATCTACT CCCATGGCTG TGCCCCCTGT
6151 GGACCTCTCT GGGCCCCTAT GACTCCTTGT GTTCTCCTTG CTCAATGCCC
6201 TGCTGAGCCC TCTGGCTCTC CCTTGCTCCC TGGACCTCTA TGTGTCTCTG
6251 TACCTCCTTG CCTCCCTTTG TTCTTGCATA TCTTTCTGAG TCCTCTGGCT
```

```
6301 CCCCCTGATT TATCCTCAGA ACTCCATCTT GTTTCAGGTT CCTGGTTCCT
6351 ATGTCCAGAC CCCTGGGCAT AGCACTGCCT GGGGATGAGA TGTTCTCATT
6401 GCTGAGAACC AGCTGAGAAG TGTTGGGTAC TTTAGACCTT TAGAGGCTGG
6451 CTTCACTAGC CTCTGGAGGT TTCTCCTCTG AGTAGCCAAT GGAGATACCC
6501 CTCCCTTGAC CCGTGGCATC AATTGGTAAA AGCCATCTAA TAATACCTAG
6551 GGCTGTTCTG AGTTCAGTCA GGCAGTAAAT AGTCATGCTG CACAGTTGAG
6601 AATATCCCCA AGAGGAGTGA GCAACCACAT CACATCCAAC CTGAGATATA
6651 TGTATAATTA GGACAGTGGT AAGAATATAA AATCGTGAAA ATATTTTTTT
6701 CACACAAAAT TTTTTTGGCT CCTGACCCTT GGACAAATTT GACCAGTTAT
6751 GACTATCAAG TTCTGTTGAA AAATACATCA CCACATGGAG AGCAAATCTC
6801 CACAGCAGGA TTGCACACTA TAATAAGAAC ATACAGCTAA GATGAAACAC
6851 ACACCTGTAG TGAAAATACA ACATTAAACT GAGAACATAC GCCATAGTAA
6901 GAACACATAA GTATCAAGAG AACACACAGC CATGGTGGGA GCCCATTGGG
6951 AGGACACACA GACAAAGTGA AATGCAGAAA GAGAGAGAGA GTGAGTGAGA
7001 GATTGTGAAA ACAGGGCCAC AGGAAACACA CAGAAATAGA GAGAGACACC
7051 AAGCCATCTA GAGATCACAG AACTTCATGG CCATGTGGCC ATAATGAGAA
7101 TGCTACTGAA CTCCTAAATG AAAAATGTCA TGTATGTTCC ATAGCTGTTG
7151 AGAGAGCCCA CAGCATGGAG AGAACACCTT ATATTAAAAA TACCCAGGCC
7201 GGGCGTGGTG AGTCACGCCT GTAATCCTAG CACTTTGGGA GGCTGAGGCA
7251 GGTGGATTGC TTGAGCGGCT TGAGCCTAGG AGTTTGAGAC CAGCCTGGGC
7301 AACATGGCAA AACCTCATCT CTACAAAAAA TATAAAAATT AGTCGGGTGT
7351 GGTAGTGCGT TCCTATAGTC CCATCTACTT CAGAGGCTGA GCCCGGAAGG
7401 TCGAGGCTTC AGTGAGCCGT GATCGTGCTA CTGCACTCCA GCCTGGGTGA
7451 CAGAGTGAGA CCATGTCTCA AAAAAAACAA AAACAAAAAA CAAAACAAAA
7501 CAAACAAACA AACAAAAAAC CCATATATAT ATATATATAC CTAGCTGAGG
7551 TGAGAATGCA CTATTTTGGT AAAATCACCA ACATGACCCA GCTACAGCAT
7601 GGGGCAGTCC CTCCCCTCTC ACTGGTAAAT TTTTCTTTCT CTGACTCACA
7651 GTTTTGTTGT TGTTGTTGCT GTTGTTTGAG ATGGAGTCTC ACTCTGTCAC
7701 CCAGGCTGGA GTGCAATGGC GCAATCTTGG TTCACTGCAA CCTCTGCCTC
7751 CTGGGTTCAA GCGATCCTCC TGCCTCAGCC TCCCGTATAG CTGGGACTAC
7801 AGGCGCATAC CACCATGCCT GGCTAATTTT TGTATTTTTT TTTGGGTTAC
7851 AATGTACTAT TTATTAATTT AATTTTTGTA TTTTTAGTAG AGATAGGGTT
7901 TCACCATGTT GGCCAGGCTG GTCTCGAACT CCTGACCTCA GGTGATCCGC
7951 CTGCCTCGGC CTCCCAAAGT GCTAGGATTA CAGGCATGAG CAACCACGCC
8001 TGGCCCCTCA TAGGTTTTTA TCTATTCTCT TTGCTTCTTC ACAACTTTGG
8051 CTTGCACGTG GACCATCATG TTCTCTCCAC TTTCTCACTA CTTCATGATC
8101 TTTCAGTCTC AGTTCCAACT GATACCTCCC TCAGTTGCTC TTTTTTCCTA
8151 GTAAGATTTC CAGAGAGGGA ATCTGAATGG CCCAGTCCAT ATTTTCAGAC
8201 CACACCACAT TAAAGTGGTT GATTGCCAGC CTATGTATTG GCTACATTAA
8251 TGGGTTGGGA ACTCATCATT TACTTCATTG CACAAAGCAG CATAGCTCTG
8301 GTTCTCAAAA TAGGGCCCCT GGGCCAGGTG TGGTGGCTCA TGCCTATAAT
8351 CCCAACACTG TGGGAGGCCG AGGGGGGCAG ATCACTTGAG TCCAGGAGTT
8401 CTAGACCAGC CTGGGCAACA TGGTGAAATC TCATCTCTAC TAAAAATACA
8451 AAAAATTAGC CAGGTGTGGT GGCATGCACC AGTAGTCCCA GCTGTTCAGG
8501 AGGCTGAGGT GGGAGGATTG CTCGAGTGTG GGAGGCAGAG ATTGCAGTGA
8551 ACCGTGACTG TGCCTCTGCA ATCCAGCCTG GGTGACAGAT TGAGACCCTG
8601 TCTCAAAAAA CAAATAAATA AAATAAAATA AATATGGTTC CTGAGCAGGG
8651 TAATTTCAGT GGGAAACCTC CCAGGGGAGG TGGATATGTC AGTCACCGCT
8701 GTATACTCAG TACACGGCTA ATAAGAGAAC TTGTGGTAGC AGCAAGAACA
8751 CTAGGTATTT ACTCAACAAA TATTTGTTGA GCATCTGATA AGAAGTGGGC
8801 ATTGTCCTAG GCACTGAGAT ACAGTAGTCA ACATGGCAGA CAAGATGCCT
8851 GCCCTGACAG GCTCTGCTAA AGTGAGAGAG GACAATAAGA AAGAGAAAGG
8901 AAGAAAGAGA ATAATTTTAG GTAATATTAA GGGTTGTAAA GAAAATAAGA
8951 CAGGATAGTG GGATAGAGGT GAGGAGAATG AGGGCTGTCT TCTGAAGAAA
9001 TGATTTTTGA GCTGAGACTT CAGTGATGAG AAGGAATTAA CCACACGATG
9051 TGCTGGAGGA AAAGCATTTT AGGGAGGGTG AGCAGCACAT ACTTCAAGGA
9101 ATCAAGAAGG AAGCCTGGTG AGGCTGGAAC ACAGAGAAAG AGCAGGTGGG
9151 TGACTTGAAA GGGCAGGGAC GGCAGTGGCC AGGTTACCTA GACCTGGTAA
9201 GGGTTTTCAA CCATAAAAGG GAGTCATCAG AAAGTCTTGA GCAGGGCTGT
9251 GATATATTCT AACTCATTTT TTATAAAAGA TCACTCTGAC TTTTTGCAGA
9301 ACATAAGTTA TAAAAGTACA AGCATGTAAG CAAGGAATCC AGCTAGCAAT
9351 CCGTGCAGTT GTCCAAATTA GAGGTGATGA CCGCTTGGAC TAGGATGATA
9401 GCAGCAGAGG TGGTGAGGAA TCACCATGAT ATATTTTGGA GGTAGAGCTG
```

FIGURE 3, page 3 of 8

9451 ACAGCATTAA CTAATAGCTA AGATAGGCCG GGTGTGGTGG CTTACGCCTG
9501 TAATCCTAGC ACTTTGGGAG GCCAAGGCGA GTGGATCACC TGAGGTCAGG
9551 AGTTCGAGAC CAGCTTGACC AACATGGTGA AACCTCGTCT CTACTAAAAA
9601 TACAAAATTA GCTGGGAATG GTGGCACATG CCTGTAATCT CAGCCTACTT
9651 GGGAGGCTGA GGCAGGAGAA TCGCTTGAAC CTGGGAGGTG AATGTTGCAG
9701 TGAGCCGAGA TTGCACCATT GCACTCCAGC CTGGGGAACA AGAGTGAAAC
9751 TCCGTCTCTA AATAAATGAA TGAATGAATG ATATCAGTCA GAGTAGGGAA
9801 GGGAAAAGAG GCTTCAAGAA TGACTCAGCT TTCGTGGACT CAGCAACTGA
9851 GTGGCTGGTG GTTTTGTTTT CTAAAATTGG GAAAGACTAG GGAGTGTGTG
9901 TGTTGGTGGG GGGCAGAAAT CAGTTTGGGC ATATTAGGTT TTGGGTGCCT
9951 ATTGGCACCC CATAAGCATG TCAGGTAGGC AGCTGATTTG GAGCCTAAAC
10001 CTCAAAGGAG AGGTCAGTCA GAGCTGACGA GAACAGATTG GAAGTCATCA
10051 GCATATAGAT GGCATTTAAA GCCCCTGGAC TAGGTGAGAT TACCAAGGAA
10101 GTGAAGGTAG AGAGAGAAGA GAAGAGGCCC AAAGTAGGGG ATTCCAATAT
10151 TTAGATATCA GGTTGAAGAA AAGAGTAGTC AAAAAAGATA AGAGGAATAC
10201 TGGGAGAGTC AGGTGTCACA GAAGCCAAGT TCCAAAAAAA GACATTTAAA
10251 GGAGAAGGAA GTAGTGAGCA GTCCAGTGCT CCTGAGAGGT AGGGTCAGAT
10301 GAGAACAGAG AATTGACCAT GAGATTTCGC AAATTGGAGA ATACTAGCAA
10351 CCTGGATAAG AACAATTTCA ATGGTTGAGG GAAACAGAAG TGTAATTGAA
10401 GAGGATTGAG GAAAAAAGAC AAATGGGAGC CTAGATAATT CCTTAATAAG
10451 TTGTTGTGAA AAGAGGAGAA GAAAAACGGG GTGCTAGCCC AGCTACTCCC
10501 TCACTCTTCC ACCACCTCAT AGGGAGAGAC TGGAGAACAC AGCCAGAGTG
10551 AGAACATTCA GTAGAAGTGG TGCTTCCTTT TTAAGTTCTG GACACTGTAT
10601 TTCATTATCT ATAACCGCAT CTCTGTACAT GGACACCTGA AATCCTTAGG
10651 GAGTGCCCGC CAACCCCATG ATGTTGGCCT TACCTGGAAA CTTAGCCACT
10701 GTTTTCCACA CTTGCCTTTC TTTCAGGCAC CTGCTGATTC CAGTTTCAGC
10751 CAGGGCACAG TGCCCAACAT TGCTGACCAA GTCTTGCTCT ATTTCTCCTT
10801 CTCACCTGGC CTCTTCCATC TTGGCCTCTG GATGCATTCT CTCCCTCTCA
10851 TGACTCATTT CTGCATTCAT CACTAGCCTC TTCTCTGCCT GGGCTTCTGC
10901 CAGCGGCCCT AGAGCAACCT ATGGTATTCC ACAGGGACCC ACTACTGGCG
10951 TCTGGACACC AGCCGGGATG GCTGGCATAG CTGGCCCATT GCTCATCAGT
11001 GGCCCCAGGG TCCTTCAGCA GTGGATGCTG CCTTTTCCTG GGAAGAAAAA
11051 CTCTATCTGG TCCAGGTGTG TATTGGGGGA GAGGCTTGAG GTAGAGACTG
11101 GGACAAGCAT ATCCAACTCT GTATTTATTA CCATCCTTTG TCCTCCAGGG
11151 CACCCAGGTA TATGTCTTCC TGACAAAGGG AGGCTATACC CTAGTAAGCG
11201 GTTATCCGAA GCGGCTGGAG AAGGAAGTCG GGACCCCTCA TGGGATTATC
11251 CTGGACTCTG TGGATGCGGC CTTTATCTGC CCTGGGTCTT CTCGGCTCCA
11301 TATCATGGCA GGTGAGGGGC TTCTGGGTGC TTAGAGGGCA GCTTGTTCTG
11351 CTACCTGTCT GTGGCATAGA TCCCCACCAG GGCATGAGAA GGCCTAGGTC
11401 AGGATCCCCA GGGCATGAGA AGGCCTAGGT CAGGATCCCC ATGACATGGA
11451 AGCCATGCTA TGTTTGGTGC CTTCTCCCCA GGACGGCGGC TGTGGTGGCT
11501 GGACCTGAAG TCAGGAGCCC AAGCCACGTG GACAGAGCTT CCTTGGCCCC
11551 ATGAGAAGGT AGACGGAGCC TTGTGTATGG AAAAGTCCCT TGGCCCTAAC
11601 TCATGTTCCG CCAATGGTCC CGGCTTGTAC CTCATCCATG GTCCCAATTT
11651 GTACTGCTAC AGTGATGTGG AGAAACTGAA TGCAGCCAAG GCCCTTCCGC
11701 AACCCCAGAA TGTGACCAGT CTCCTGGGCT GCACTCACTG AGGGGCCTTC
11751 TGACATGAGT CTGGCCTGGC CCCACCTCCT AGTTCCTCAT AATAAAGACA
11801 GATTGCTTCT TCGCTTCTCA CTGAGGGGCC TTCTGACATG AGTCTGGCCT
11851 GGCCCCACCT CCCCAGTTTC TCATAATAAA GACAGATTGC TTCTTCACTT
11901 GAATCAAGGG ACCTTGGTCG TGAAACAATC TTCTTTCTTT GAGTTGAAAA
11951 GTTAGCACTT CTCCTTTGAG GGTGTCGAGC TCAAACAAGG CTGTGAGAAA
12001 CAAGGGAGGG GAGCACTAAG GGGCAAACCT ATCTCTGCGC AGATGATTCT
12051 TAGGTCCAGA TCATAAACTA GCTCTTTGCA GACTATCTAC ACATAGTGGG
12101 GGGAAAGAGA ACCAGAGTCG GAAGAGGAAC AGCTGAGTTT ATACAGCAAG
12151 TAAGAGGTGG AGCTAGGACT CTGATTCAAC TTGCTGGTAG ATGGCCACAA
12201 CCCAGCCGCA AGGCATCAGA AACAACAGGG CCTGGGGCAA CTATGCATGT
12251 GCAAAGAGGA TTGGCTCAGA GTTGTGGGGT AGGAGGTCCA ATCTGGGGGA
12301 CCTCAAATTA TGGTTCTGGG TGATTCAAGT AACACCACTC ATGGCTTGTG
12351 TTGCCATGAG TTAGGCATGA CAAGTGGAAT GAAGTTGAAG TGGGGAAACA
12401 GAAATACACC AGCTGTGTGT CAGAGGCAAG CTGGAGAGAG AGAAGAAAGA
12451 ATGAATGGCA CCATGGAGCA CATTTGCAGA ACACAGTCCC TGGGAGTCTT
12501 GCTGGAGCCT CAGGAGCTTT GCTGGCACAG AGGATCTGGC CTACCCAATT
12551 AGCCTCCTGG GTATCTGCAC CATCTAGACC AGCAAATGTC ACTGGCAAGG

FIGURE 3, page 4 of 8

```
12601 AGGTTGCAGT GCTTGGTTAT TTTCTGGTCA TAAACTGGTG AAGGCTTTGG
12651 GTTCCAAATT TGCTGACAGC TGTTTAACTG GGAATTGGGC CTAGACTATA
12701 GGTAGCTATG TCTCAGACAA GGCCCTATTC CTCCACTGCC TTTACAACCC
12751 AGCTGAGGTT GGAGGCTGGC TTGTTTCAGC CTCAAAAAAT AGCCTGAGTT
12801 TCCAGCAGAG GGCCCTTATT CTGAGCTTCC GTGTCCTAGC CTCATTTTCC
12851 TTTCCTGTAA AATAGACACA ATGCCACCCA CCTTCCAGTG ACAATGAATA
12901 TAGACTCAAA CCCATCCCTT GAACTGTCTT GGGAAGGGGC TCTGGACGTA
12951 GACCCAGACT GTGGCTCATG GCCTCATGTG ATCTGGAGTC AGCCCCTCCC
13001 AACCTGTCAG CCATTTGCTC CGTAGGACTT TGATGGGTAG AGTAGTAGCT
13051 AACAAGCTCT GACTGTCACA CAAGGCTTTG TACTGGGAGG CCAGGCTATA
13101 GAGTGGCTCC AGCTTAAAGG GCTGGGAGCT GGGGGACAGT GTCTCAGATT
13151 AGGGTCTAAC TAGGAAGTTG ACTGGAGCTG AGAACAGAGG TTAGGGGCCA
13201 AGCAGCAGGG TTGTGGGTCT ACTCCTTAGG AGCACCTTGA GCTTTACTTT
13251 TCATTCCTAA TGGTGTCTTG GATGGCTACC CTCACGGGGT TGGCTGCTAG
13301 TCTAAGGGGT GGAGACAAGG ACAGAGTTTC AGGTCTGGTC CTTATCAAGT
13351 TCATGCACTA CACTTGGGAC CACTGCTGCA TCATGCCAGG GAGCCTAGAG
13401 GTGTCTAAAC AGTTATCCAA CAACTGTGAT ACCCAAGGTT AACTTTCTCT
13451 TGTTTTCAGA GGCAGGGAGT ACTAAGTCTC CCCTTTCTCC TTTCCTCCCA
13501 CGTGTTCTCT TGCAGGGAAT CCTCTAGCTT GTCTCCAGGG AACTCCCAGA
13551 AATGGTTTGT TTCAGTCAGT TTAGGCTGCT ATAAGAGAAT ATCTTAGAGT
13601 GGGTAATCTA TCAGCAATAG GAATTTATTG TTCACAATTC TGGAGGCTGG
13651 AAAATCCAAG ATCAAGGCTC CAGCAGGTTC AGTGTCTGCT GAGTGCTTGT
13701 TCTGCTTCGA AGATGGCACC TTTTTGCTGT GTTCTCA  (SEQ ID NO:3)
```

FEATURES:

Genewise results:

```
Start:   2001
Exon:    2001-2083
Intron:  2084-2233
Exon:    2234-2292
Intron:  2293-2413
Exon:    2414-2485
Intron:  2486-2665
Exon:    2666-2787
Intron:  2788-4442
Exon:    4443-4596
Intron:  4597-5774
Exon:    5775-5906
Intron:  5907-10934
Exon:    10935-11065
Intron:  11066-11148
Exon:    11149-11311
Intron:  11312-11481
Exon:    11482-11738
Stop:    11739
```

Sim4 results:

```
Exon:    1987-2083, (Transcript Position: 1-97)
Exon:    2234-2292, (Transcript Position: 98-156)
Exon:    2414-2485, (Transcript Position: 157-228)
Exon:    2666-2787, (Transcript Position: 229-350)
Exon:    4443-4596, (Transcript Position: 351-504)
Exon:    5775-5906, (Transcript Position: 505-636)
Exon:    10935-11065, (Transcript Position: 637-767)
Exon:    11149-11311, (Transcript Position: 768-930)
Exon:    11482-13737, (Transcript Position: 931-3186)
```

CHROMOSOME MAP POSITION:

Chromosome 11

ALLELIC VARIANTS (SNPs):

DNA Position	Major	Minor	Domain
1106	C	T	Intron
4344	A	G	Intron
7078	T	A	Intron
10841	C	G	Intron
10850	A	G	Intron
12727	G	A	Exon, 3' UTR
13164	T	G	Exon, 3' UTR
13285	T	C	Exon, 3' UTR
13654	A	G	Exon, 3' UTR
13699	G	C	Exon, 3' UTR

Context:

```
DNA
Position
1106      AATTCCTAATATTCTCCTACCCTAGAGGGGGGGAAATTGTCAGAAATTTGGCTGCAACAC
          TAGCAACACTACTCAGTACTTGAAATGCATTTTTGCATTTTTTTCATTCAACAAATATTT
          CTGGAACAACTCTTATATGCCAGGCACTATTTTAGGAGTCAGGGATATATAATGGTAAAC
          AAGACAGGCAAAACAAAGCAAAGCAACAACAACCATCACCAGATAAGTAGACAGATGAAA
          GAATTTCAAGTTTTAGTAAGTAAAATAAAACAAGCAAGGGTCTGAAATGGCTAGATAAGG
          [C,T]
          GGTCAAGAAAGGCTTCATTGAGAAGGTAGCATTTAAGCAGGAGTCAGCTAGAAATATTGT
          GAAATTCCAGTTACAGTTCTATTTGTTCTGGGTTGGTTAAATAAAGCTTTTTCCCCCAAG
          GTGGAAACTACCAAGAAAGACTAATTACTAGTAGTGGTGGTGCTCTCTGGAAGAGAGACA
          CCTCCTGTTTCTGCCTCATTACTGTCAACCCTTCACTTCCAGGCACTTTTTGCAAAGCCC
          TTTGCCAGTCAGGGAAGGCGAGAGGCTGGGCATGGGGCTTGGACATTTGACAACAGTGAG

4344      TGCCTGGCATATAGTAGTTATTCAGTGACTGTTTCTTGGATGAATGAATGAATGAATAAA
          TAAATGAAGAAATGAATGAAGAAATAACGTATGGGTGATTGCAGGATGAACAGTTGTGGA
          TATGTTTGTCAACACTGATAGTGTTGCAGATAAATGTGCCACAGGAGTGTCTGGGTACAG
          AGCTAGAGGCATGTGTGTTATAGTAATAGTGACTGGATTTGCACAAACTGAGAGTGTGTA
          ATGTGCAAAAGGACAGCACATTGTTGTCCACAGATGGACTGAGAATGTGTAGGGCCACAG
          [A,G]
          AGGATATCGTATAAGCACAGTAGATAAAAAATGTGTGTAAATGCAGAGTGGCAGTATCTG
          GGGATGCACAGTCAAAAAGAGAGTACTTTTGAATGCAGGGGGACAAAGTCTGGGTATACC
          CTCCTGAAAAGAAGGAGAAAGGATACCCAAAGTTGCTCCAAGATGAATTTCCTGGAATCC
          CATCCCCACTGGATGCAGCTGTGGAATGTCACCGTGGAGAATGTCAAGCTGAAGGCGTCC
          TCTTCTTCCAAGGTCAGTCCAGGCTGGAATCCAAGAACCTGGAGTAGTGGTGGGTTGGTA

7078      TCACCACATGGAGAGCAAATCTCCACAGCAGGATTGCACACTATAATAAGAACATACAGC
          TAAGATGAAACACACACCTGTAGTGAAAATACAACATTAAACTGAGAACATACGCCATAG
          TAAGAACACATAAGTATCAAGAGAACACACAGCCATGGTGGGAGCCCATTGGGAGGACAC
          ACAGACAAAGTGAAATGCAGAAAGAGAGAGAGAGTGAGTGAGAGATTGTGAAAACAGGGC
          CACAGGAAACACACAGAAATAGAGAGAGACACCAAGCCATCTAGAGATCACAGAACTTCA
          [T,A]
          GGCCATGTGGCCATAATGAGAATGCTACTGAACTCCTAAATGAAAAATGTCATGTATGTT
          CCATAGCTGTTGAGAGAGCCCACAGCATGGAGAGAACACCTTATATTAAAAATACCCAGG
          CCGGGCGTGGTGAGTCACGCCTGTAATCCTAGCACTTTGGGAGGCTGAGGCAGGTGGATT
          GCTTGAGCGGCTTGAGCCTAGGAGTTTGAGACCAGCCTGGGCAACATGGCAAAACCTCAT
          CTCTACAAAAAATATAAAAATTAGTCGGGTGTGGTAGTGCGTTCCTATAGTCCCATCTAC

10841     AGCCAGAGTGAGAACATTCAGTAGAAGTGGTGCTTCCTTTTTAAGTTCTGGACACTGTAT
          TTCATTATCTATAACCGCATCTCTGTACATGGACACCTGAAATCCTTAGGGAGTGCCCGC
          CAACCCCATGATGTTGGCCTTACCTGGAAACTTAGCCACTGTTTTCCACACTTGCCTTTC
          TTTCAGGCACCTGCTGATTCCAGTTTCAGCCAGGGCACAGTGCCCAACATTGCTGACCAA
          GTCTTGCTCTATTTCTCCTTCTCACCTGGCCTCTTCCATCTTGGCCTCTGGATGCATTCT
          [C,G]
          TCCCTCTCATGACTCATTTCTGCATTCATCACTAGCCTCTTCTCTGCCTGGGCTTCTGCC
          AGCGGCCCTAGAGCAACCTATGGTATTCCACAGGGACCCACTACTGGCGTCTGGACACCA
          GCCGGGATGGCTGGCATAGCTGGCCCATTGCTCATCAGTGGCCCCAGGGTCCTTCAGCAG
```

TGGATGCTGCCTTTTCCTGGGAAGAAAAACTCTATCTGGTCCAGGTGTGTATTGGGGGAG
AGGCTTGAGGTAGAGACTGGGACAAGCATATCCAACTCTGTATTTATTACCATCCTTTGT

10850 GAGAACATTCAGTAGAAGTGGTGCTTCCTTTTTAAGTTCTGGACACTGTATTTCATTATC
TATAACCGCATCTCTGTACATGGACACCTGAAATCCTTAGGGAGTGCCCGCCAACCCCAT
GATGTTGGCCTTACCTGGAAACTTAGCCACTGTTTTCCACACTTGCCTTTCTTTCAGGCA
CCTGCTGATTCCAGTTTCAGCCAGGGCACAGTGCCCAACATTGCTGACCAAGTCTTGCTC
TATTTCTCCTTCTCACCTGGCCTCTTCCATCTTGGCCTCTGGATGCATTCTCTCCCTCTC
[A,G]
TGACTCATTTCTGCATTCATCACTAGCCTCTTCTCTGCCTGGGCTTCTGCCAGCGGCCCT
AGAGCAACCTATGGTATTCCACAGGGACCCACTACTGGCGTCTGGACACCAGCCGGGATG
GCTGGCATAGCTGGCCCATTGCTCATCAGTGGCCCCAGGGTCCTTCAGCAGTGGATGCTG
CCTTTTCCTGGGAAGAAAAACTCTATCTGGTCCAGGTGTGTATTGGGGGAGAGGCTTGAG
GTAGAGACTGGGACAAGCATATCCAACTCTGTATTTATTACCATCCTTTGTCCTCCAGGG

12727 CAAGCTGGAGAGAGAGAAGAAAGAATGAATGGCACCATGGAGCACATTTGCAGAACACAG
TCCCTGGGAGTCTTGCTGGAGCCTCAGGAGCTTTGCTGGCACAGAGGATCTGGCCTACCC
AATTAGCCTCCTGGGTATCTGCACCATCTAGACCAGCAAATGTCACTGGCAAGGAGGTTG
CAGTGCTTGGTTATTTTCTGGTCATAAACTGGTGAAGGCTTTGGGTTCCAAATTTGCTGA
CAGCTGTTTAACTGGGAATTGGGCCTAGACTATAGGTAGCTATGTCTCAGACAAGGCCCT
[G,A]
TTCCTCCACTGCCTTTACAACCCAGCTGAGGTTGGAGGCTGGCTTGTTTCAGCCTCAAAA
AATAGCCTGAGTTTCCAGCAGAGGGCCCTTATTCTGAGCTTCCGTGTCCTAGCCTCATTT
TCCTTTCCTGTAAAATAGACACAATGCCACCCACCTTCCAGTGACAATGAATATAGACTC
AAACCCATCCCTTGAACTGTCTTGGGAAGGGGCTCTGGACGTAGACCCAGACTGTGGCTC
ATGGCCTCATGTGATCTGGAGTCAGCCCCTCCCAACCTGTCAGCCATTTGCTCCGTAGGA

13164 AGACACAATGCCACCCACCTTCCAGTGACAATGAATATAGACTCAAACCCATCCCTTGAA
CTGTCTTGGGAAGGGGCTCTGGACGTAGACCCAGACTGTGGCTCATGGCCTCATGTGATC
TGGAGTCAGCCCCTCCCAACCTGTCAGCCATTTGCTCCGTAGGACTTTGATGGGTAGAGT
AGTAGCTAACAAGCTCTGACTGTCACACAAGGCTTTGTACTGGGAGGCCAGGCTATAGAG
TGGCTCCAGCTTAAAGGGCTGGGAGCTGGGGGACAGTGTCTCAGATTAGGGTCTAACTAG
[T,G]
AAGTTGACTGGAGCTGAGAACAGAGGTTAGGGGCCAAGCAGCAGGGTTGTGGGTCTACTC
CTTAGGAGCACCTTGAGCTTTACTTTTCATTCCTAATGGTGTCTTGGATGGCTACCCTCA
CGGGGTTGGCTGCTAGTCTAAGGGGTGGAGACAAGGACAGAGTTTCAGGTCTGGTCCTTA
TCAAGTTCATGCACTACACTTGGGACCACTGCTGCATCATGCCAGGGAGCCTAGAGGTGT
CTAAACAGTTATCCAACAACTGTGATACCCAAGGTTAACTTTCTCTTGTTTTCAGAGGCA

13285 GGAGTCAGCCCCTCCCAACCTGTCAGCCATTTGCTCCGTAGGACTTTGATGGGTAGAGTA
GTAGCTAACAAGCTCTGACTGTCACACAAGGCTTTGTACTGGGAGGCCAGGCTATAGAGT
GGCTCCAGCTTAAAGGGCTGGGAGCTGGGGGACAGTGTCTCAGATTAGGGTCTAACTAGG
AAGTTGACTGGAGCTGAGAACAGAGGTTAGGGGCCAAGCAGCAGGGTTGTGGGTCTACTC
CTTAGGAGCACCTTGAGCTTTACTTTTCATTCCTAATGGTGTCTTGGATGGCTACCCTCA
[T,C]
GGGGTTGGCTGCTAGTCTAAGGGGTGGAGACAAGGACAGAGTTTCAGGTCTGGTCCTTAT
CAAGTTCATGCACTACACTTGGGACCACTGCTGCATCATGCCAGGGAGCCTAGAGGTGTC
TAAACAGTTATCCAACAACTGTGATACCCAAGGTTAACTTTCTCTTGTTTTCAGAGGCAG
GGAGTACTAAGTCTCCCCTTTCTCCTTTCCTCCCACGTGTTCTCTTGCAGGGAATCCTCT
AGCTTGTCTCCAGGGAACTCCCAGAAATGGTTTGTTTCAGTCAGTTTAGGCTGCTATAAG

13654 TGCACTACACTTGGGACCACTGCTGCATCATGCCAGGGAGCCTAGAGGTGTCTAAACAGT
TATCCAACAACTGTGATACCCAAGGTTAACTTTCTCTTGTTTTCAGAGGCAGGGAGTACT
AAGTCTCCCCTTTCTCCTTTCCTCCCACGTGTTCTCTTGCAGGGAATCCTCTAGCTTGTC
TCCAGGGAACTCCCAGAAATGGTTTGTTTCAGTCAGTTTAGGCTGCTATAAGAGAATATC
TTAGAGTGGGTAATCTATCAGCAATAGGAATTTATTGTTCACAATTCTGGAGGCTGGAAA
[A,G]
TCCAAGATCAAGGCTCCAGCAGGTTCAGTGTCTGCTGAGTGCTTGTTCTGCTTCGAAGAT
GGCACCTTTTTGCTGTGTTCTCA

13699 AGGTGTCTAAACAGTTATCCAACAACTGTGATACCCAAGGTTAACTTTCTCTTGTTTTCA
GAGGCAGGGAGTACTAAGTCTCCCCTTTCTCCTTTCCTCCCACGTGTTCTCTTGCAGGGA
ATCCTCTAGCTTGTCTCCAGGGAACTCCCAGAAATGGTTTGTTTCAGTCAGTTTAGGCTG

CTATAAGAGAATATCTTAGAGTGGGTAATCTATCAGCAATAGGAATTTATTGTTCACAAT
TCTGGAGGCTGGAAAATCCAAGATCAAGGCTCCAGCAGGTTCAGTGTCTGCTGAGTGCTT
[G,C]
TTCTGCTTCGAAGATGGCACCTTTTTGCTGTGTTCTCA